

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 29, 2003, 13:56:00 ; Search time 26 Seconds  
(without alignments)  
576.131 Million cell updates/sec

Title: US-09-818-143-22  
Perfect score: 539  
Sequence: 1 MEALNVNFPKSRFPEDAG.....RLKVGPRQASQRLKETGLC 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters: 566894

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/FCI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539	100.0	99	9	US-09-818-143-22
2	539	100.0	99	10	US-09-974-298-149
3	79.5	14.7	322	9	US-09-789-836-6
4	79.5	14.7	322	12	US-09-789-831-6
5	79.5	14.7	345	9	US-09-789-836-5
6	79.5	14.7	345	12	US-09-789-831-5
7	79.5	14.7	345	12	US-10-251-661-10
8	73.5	13.6	345	15	US-10-060-036-183
9	70.5	13.1	419	16	US-10-138-098-25
10	69.5	12.9	419	16	US-10-138-098-27
11	69	12.8	323	15	US-10-205-823-62
12	69	12.8	323	15	US-10-177-293-53
13	69	12.8	4123	15	US-10-213-509-5
14	67.5	12.5	627	12	US-10-349-836-6
15	67	12.4	597	9	US-09-853-386-93

Sequence 22, Appl  
Sequence 149, Appl  
Sequence 6, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 10, Appl  
Sequence 183, Appl  
Sequence 25, Appl  
Sequence 27, Appl  
Sequence 62, Appl  
Sequence 53, Appl  
Sequence 5, Appl  
Sequence 6, Appl  
Sequence 93, Appl

Sequence 115, Appl  
Sequence 6, Appl  
Sequence 1316, Appl  
Sequence 863, Appl  
Sequence 12, Appl  
Sequence 58, Appl  
Sequence 8, Appl  
Sequence 8, Appl  
Sequence 102, Appl  
Sequence 61, Appl  
Sequence 61, Appl  
Sequence 13173, A  
Sequence 20, Appl  
Sequence 36, Appl  
Sequence 14, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 827, Appl  
Sequence 30, Appl  
Sequence 5611, Appl  
Sequence 103, Appl  
Sequence 94, Appl  
Sequence 206, Appl  
Sequence 12, Appl  
Sequence 20, Appl  
Sequence 18, Appl  
Sequence 4323, Appl  
Sequence 1546, Appl  
Sequence 198, Appl  
Sequence 198, Appl

## ALIGNMENTS

RESULT 1

US-09-818-143-22  
Sequence 22, Application US/09818143  
Patent No. US20020019000A1  
GENERAL INFORMATION:  
APPLICANT: Walker, Michael G.  
APPLICANT: Volkmut, Wayne  
APPLICANT: Klingler, Tod M.  
TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES  
FILE REFERENCE: PB-0004 CIP  
CURRENT APPLICATION NUMBER: US/09/818.143  
CURRENT FILING DATE: 2001-03-26  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PERL Program  
SEQ ID NO 22  
LENGTH: 99  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE: -  
OTHER INFORMATION: 1556751CD1  
US-09-818-143-22

Query Match 100.0%; Score 539; DB 9; Length 99;  
Best Local Similarity 100.0%; Pred. No. 4.4e-48;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MEALNVNFPKSRFPEDAGKESGGGFCVPAARPTMTGTCSPGLNFSFQRKEN 60

Db 1 MEALNVNFPKSRFPEDAGKESGGGFCVPAARPTMTGTCSPGLNFSFQRKEN 60

Oy 61 RACACQWAGPAPKPMCMVRLKVGPRQASQRLKETGLC 99

Db 61 RACACQWAGPAPKPMCMVRLKVGPRQASQRLKETGLC 99

RESULT 2

US-09-974-298-149

; Sequence 149, Application US/09974298

```
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 149
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1556751CD1
; US-09-974-298-149

Query Match      100.0%; Score 539; DB 10; Length 99;
Best Local Similarity 100.0%; Pred. No. 4.4e-48;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALANVFPKSFPRPEDAGKESGSGGFCVPAARPOTMTGPGSCSPGLQNFSPORKEN 60
Db 1 MEALANVFPKSFPRPEDAGKESGSGGFCVPAARPOTMTGPGSCSPGLQNFSPORKEN 60
QY 61 RACACWQAGAPKPNMVCRLKVGPRPOASORKKETGLC 99
Db 61 RACACWQAGAPKPNMVCRLKVGPRPOASORKKETGLC 99

RESULT 3
US-09-789-836-6
; Sequence 6, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
; APPLICANT: BRIGHAM, KENNETH L.
; APPLICANT: STECENKO, ARLENE A.
; APPLICANT: SEALY, LINDA
; TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
; FILE REFERENCE: N-6977
; CURRENT APPLICATION NUMBER: US/09/789,836
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/183,584
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-789-836-6

Query Match      14.7%; Score 79.5; DB 9; Length 322;
Best Local Similarity 26.2%; Pred. No. 2.5;
Matches 32; Conservative 10; Mismatches 39; Indels 41; Gaps 5;

QY 10 PRKSFPRPEDAGKESGSGGFCVPAARPOTMTGPGSCSPGLQNFSP 55
Db 158 PADCKRKEEAGAGGGAGMAAGFPYALRAYLGYQAVPSGSGSLSTSSSSPPGTP--SP 215
QY 56 QRKENRACACWQAGAP-----CVPAAARPOTMTGPGSCSPGLQNFSP 55
Db 216 ADAPKAPTACYAGAGPAPSOVKSKAKKTVDKHSDEYKIRERNIAVR--KSRDKAKMRN 273
QY 93 LK 94
Db 274 LE 275

US-09-789-831-6
; Sequence 6, Application US/09789831
; Patent No. US20030166586A1
; GENERAL INFORMATION:
; APPLICANT: SEALY, LINDA
; TITLE OF INVENTION: C/EBP-BETA ISOFORMS AND METHODS OF USE IN CELL REGULATION
; FILE REFERENCE: AND ANTI-TUMORIGENESIS
; FILE REFERENCE: N-6978
; CURRENT APPLICATION NUMBER: US/09/789,831
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/183,532
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-789-831-6

Query Match      14.7%; Score 79.5; DB 12; Length 322;
Best Local Similarity 26.2%; Pred. No. 2.5;
Matches 32; Conservative 10; Mismatches 39; Indels 41; Gaps 5;

QY 10 PRKSFPRPEDAGKESGSGGFCVPAARPOTMTGPGSCSPGLQNFSP 55
Db 158 PADCKRKEEAGAGGGAGMAAGFPYALRAYLGYQAVPSGSGSLSTSSSSPPGTP--SP 215
QY 56 QRKENRACACWQAGAP-----CVPAAARPOTMTGPGSCSPGLQNFSP 55
Db 216 ADAPKAPTACYAGAGPAPSOVKSKAKKTVDKHSDEYKIRERNIAVR--KSRDKAKMRN 273
QY 93 LK 94
Db 274 LE 275

US-09-789-836-5
; Sequence 5, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
; APPLICANT: BRIGHAM, KENNETH L.
; APPLICANT: STECENKO, ARLENE A.
; APPLICANT: SEALY, LINDA
; TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
; FILE REFERENCE: N-6977
; CURRENT APPLICATION NUMBER: US/09/789,836
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/183,584
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-789-836-5

Query Match      14.7%; Score 79.5; DB 9; Length 345;
Best Local Similarity 26.2%; Pred. No. 2.7;
Matches 32; Conservative 10; Mismatches 39; Indels 41; Gaps 5;

QY 10 PRKSFPRPEDAGKESGSGGFCVPAARPOTMTGPGSCSPGLQNFSP 55
Db 181 PADCKRKEEAGAGGGAGMAAGFPYALRAYLGYQAVPSGSGSLSTSSSSPPGTP--SP 238
QY 56 QRKENRACACWQAGAP-----CVPAAARPOTMTGPGSCSPGLQNFSP 55
Db 239 ADAPKAPTACYAGAGPAPSOVKSKAKKTVDKHSDEYKIRERNIAVR--KSRDKAKMRN 296
QY 93 LK 94
Db 274 LE 275
```

```

RESULT 2
US-09-252-991A-32747
; Sequence 32747, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,789
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

```



Result No.	Score	Query		Length	DB	ID	Description
		Match	Length				
1	539	100.0	99	21	AA070744		P50-2 protein enc
2	539	100.0	99	23	AA014785		Human matrix-remod
3	539	100.0	99	24	ABU57648		Differentially exp
4	80	14.8	115	22	AAU14145		Human novel protei
5	80	14.8	884	22	AB969232		Drosophila melanog
6	79.5	14.7	322	22	AA093702		Human CCAAT enhanc
7	79.5	14.7	322	22	AAE11945		Human CCAAT/enhanc
8	79.5	14.7	345	12	AA014608		Nuclear factor C/E
9	79.5	14.7	345	12	AAU97001		Human CCAAT enhanc

PD 20-APR-2000.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

XX PF 06-OCT-1999; 99WO-US23315.  
 XX PR 09-OCT-1998; 98US-0169289.  
 XX PA (INCY-) INCYTE PHARM INC.  
 XX PI Walker MG, Volkmanth W, Klingler TM;  
 XX DR WPI: 2000-317934/27.  
 XX DR N-PSDB; AAZ52354.  
 XX PT Protein co-expressed with matrix-remodeling proteins. Useful in the  
 PT diagnosis and treatment of cancer, cardiomyopathy, arthritis,  
 PT angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and  
 PT ulceration.  
 XX PS Claim 1; Page 53-54; 55pp; English.  
 XX CC The present sequence is PSEQ protein encoded by NSEQ gene that is  
 C co-expressed with one or more known matrix-remodeling genes in a number  
 CC of biological samples using an expression vector. The gene, protein, and  
 CC antibody sequences can be used in the diagnosis, and treatment or  
 CC prevention of a disease associated with its altered expression. The  
 CC diseases that can be treated are matrix-remodeling diseases, including  
 CC cancer, cardiomyopathy, arthritis, angiogenesis, diabetic necrosis,  
 CC atherosclerosis, fibrosis, and ulceration.  
 XX SQ Sequence 99 AA;  
 Query Match 100.0%; Score 539; DB 21; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-51;  
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEALANVFPKSRFPEDAGKESGGGFCVPAARPTMTGTPSCSPGLQNFSPQRKEN 60  
 Db 1 MEALANVFPKSRFPEDAGKESGGGFCVPAARPTMTGTPSCSPGLQNFSPQRKEN 60  
 QY 61 RACACWQAGPAPKPMCVRLKVGPRQASQKRLKETGLC 99  
 Db 61 RACACWQAGPAPKPMCVRLKVGPRQASQKRLKETGLC 99  
 RESULT 2  
 AAO14785  
 ID AAO14785 standard; Protein; 99 AA.  
 XX AC AAO14785;  
 X DT 11-JUL-2002 (first entry)  
 XX DE Human matrix-remodeling-associated protein 2.  
 DE Human; matrix-remodeling gene; extracellular matrix;  
 KW matrix-remodeling-associated nucleotide; screening;  
 KW matrix remodeling-associated disease; angiogenesis; arthritis;  
 KW atherosclerosis; cancer; cardiomyopathy; diabetic necrosis; fibrosis;  
 KW ulceration.  
 XX OS Homo sapiens.  
 OS US2002019000-A1.  
 PN 14-FEB-2002.  
 PD 26-MAR-2001; 2001US-0818143.  
 XX PF 09-OCT-1998; 98US-0169289.  
 XX PR (WALK/) WALKER M G.  
 XX PA (VOLK/) VOLKMANTH W.  
 XX PA (KLIN/) KLINGLER T M.

PI Walker MG, Volkmanth W, Klingler TM;  
 XX WPI: 2002-338319/37.  
 XX DR N-PSDB; AAL42450.  
 XX PT New isolated polynucleotide coexpressed with matrix-remodeling genes,  
 PT useful in diagnosis, prognosis, prevention and treatment of diseases,  
 PT associated with matrix-remodeling such as angiogenesis, arthritis and  
 PT cancer.  
 XX PS Examples; Fig 2; 63pp; English.  
 XX CC The invention comprises human nucleotide sequences which are co-expressed  
 CC with matrix-remodeling genes. Matrix-remodeling is associated with the  
 CC construction, destruction and reorganization of extracellular matrix  
 CC components. The matrix-remodeling-associated nucleotides of the invention  
 CC are useful for screening for and purifying ligands that specifically bind  
 CC to the nucleotides of the invention. The matrix-remodeling-associated  
 CC nucleotides of the invention are also useful in the diagnosis, prognosis,  
 CC prevention, treatment and evaluation of therapies for diseases associated  
 CC with matrix remodeling (e.g., angiogenesis, arthritis, atherosclerosis,  
 CC cancer, cardiomyopathy, diabetic necrosis, fibrosis and ulceration). The  
 CC present amino acid sequence is encoded by a human matrix-remodeling-  
 CC associated nucleotide of the invention.  
 XX SQ Sequence 99 AA;  
 Query Match 100.0%; Score 539; DB 23; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-51;  
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEALANVFPKSRFPEDAGKESGGGFCVPAARPTMTGTPSCSPGLQNFSPQRKEN 60  
 Db 1 MEALANVFPKSRFPEDAGKESGGGFCVPAARPTMTGTPSCSPGLQNFSPQRKEN 60  
 QY 61 RACACWQAGPAPKPMCVRLKVGPRQASQKRLKETGLC 99  
 Db 61 RACACWQAGPAPKPMCVRLKVGPRQASQKRLKETGLC 99  
 RESULT 3  
 ABUS7648  
 ID ABUS7648 standard; Protein; 99 AA.  
 XX AC ABUS7648;  
 XX DT 09-APR-2003 (first entry)  
 XX DE Differentially expressed breast cancer associated protein #35.  
 DE Breast cancer; differential gene expression; BC-cDNA;  
 KW breast cancer diagnosis; breast cancer monitoring;  
 KW breast cancer treatment; breast cancer staging.  
 XX OS Homo sapiens.  
 OS US2002156263-A1.  
 PN 24-OCT-2002.  
 PD 04-OCT-2001; 2001US-0974298.  
 XX PF 05-OCT-2000; 2000US-238331P.  
 XX PR (CHEN/) CHEN H.  
 XX PA Chen H;  
 XX PI WPI: 2003-182653/18.  
 XX PT New cDNAs, which are differentially expressed in (metastatic) breast  
 PT cancer useful for diagnosing or staging, breast cancer, or for  
 PT monitoring the treatment of breast cancer in an individual.

Result No.	Query			ID	Description
	Score	Match	Length		
1	539	100.0	99	4 Q9UFS0	Q9ufs0 homo sapien
2	86.5	16.0	557	3 Q5U094	Q5u094 schizosacch
3	83.5	15.5	1196	5 Q8T433	Q8t433 drosophila
4	78	14.5	504	5 Q8S2L2	Q8s2l2 drosophila
5	78	14.5	504	5 Q8MR78	Q8mry8 drosophila
6	76.5	14.2	208	13 Q93592	Q93592 astyanax fa
7	75.5	14.0	1534	5 Q5VEH3	Q5veh3 drosophila
8	75	13.9	712	4 Q8NCN4	Q8ncn4 homo sapien
9	74.5	13.8	268	4 Q96M64	Q96m64 homo sapien
10	73.5	13.6	345	4 Q96IH2	Q96ih2 homo sapien
11	73	13.5	386	10 Q8SF2F1	Q8sf2f1 oryza sativ
12	73	13.5	945	5 Q9NRK7	Q9nkr7 teishmania
13	73	13.5	2087	5 Q8MXL2	Q8mxl2 teishmania
14	72.5	13.5	323	4 Q8TD94	Q8td94 homo sapien
15	72.5	13.5	504	5 P90730	P90730 ceratitlis c
16	72	13.4	2533	5 P90589	P90589 paramacium

```

RESULT 2
Q9U094 ID Q9U094 PRELIMINARY; PRT; 557 AA.
AC Q9U094;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 64.6 kDa protein.
GN SPC830.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
A McDougall R.M., Rajandream M.A., Bartell B.G., Ransperger U., Pohl T.;
L Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL109830; CAB52878.1; -
DR GenDB: SPombe; SPC830.05C; -
KW Hypothetical protein.
SQ SEQUENCE 557 AA; 64627 MW; 884CE3D8559EA724 CRC64;

Query Match 16.0%; Score 86.5; DB 3; Length 557;
Best Local Similarity 30.4%; Pred. No. 0.12;
Matches 28; Conservative 18; Mismatches 33; Indels 13; Gaps 6;

QY 6 NVNPKRSFR---PEDAGKES--GSGGFCVPAARQPMVTPG-----SC 45
Db 402 HTNYPDSFYNIPIYSGRETKQSHNQLSPSTPLSDNGFTYSTPHSLSNFNTCD 461

QY 54 SPQKRNACACWQNAAGPAPKPMCVRLKVGPR 85
Db 462 SLSPSSNNLYGYSTL-LHPRNPICVRQIGR 492

RESULT 3
Q8T433 ID Q8T433 PRELIMINARY; PRT; 1196 AA.
AC Q8T433;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE AT19506p.
GN CG14322.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY089375; AAL90113.1; -
DR Flybase; FBgn0038532; CG14322.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR Repeat; WD repeat.
SQ SEQUENCE 1196 AA; 133061 MW; 9ADBA0269C8DEB47 CRC64;

Query Match 15.5%; Score 83.5; DB 5; Length 1196;
Best Local Similarity 30.1%; Pred. No. 0.62;
Matches 28; Conservative 16; Mismatches 32; Indels 17; Gaps 4;

QY 2 EALANVPRK-SFRPEDAGKES---GSGGFCVPAARQPMVTPG-----SC 45
Db 481 QKLANLLSPKKTAFSTDECTCKASTLCPNQHGFIPTAQDSSTVSPCKQPPSSPPSQSP 540

QY 46 SSPGLQNFSPQR-KENRACACWQNAAGPAPKPM 77
Db 541 SSATAQSSPMALQKQSCPAQSCSTPIQENPM 573

RESULT 4
Q8SZL2 ID Q8SZL2 PRELIMINARY; PRT; 504 AA.
AC Q8SZL2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE RH16046p.
GN CYP6A21 OR CG10247.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AY070678; AAL48149.1; -
DR Flybase; FBgn0033981; Cyp6a21.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 504 AA; 57842 MW; 32DA87A519DF557A CRC64;

Query Match 14.5%; Score 78; DB 5; Length 504;
Best Local Similarity 30.5%; Pred. No. 1.1;
Matches 18; Conservative 11; Mismatches 26; Indels 4; Gaps 2;

QY 30 CVPAAQPQMTVTPGSCSPQLQNFSPQKRNACACWQNAAGPAPKPMCVRLKVGPRQA 88
Db 405 CGAMHRDEKLYANPNTFNP--DNFSPERVKERDVSVMPLFGDGPRN--CIGMRFGQMQA 459

RESULT 5
Q8MY8 ID Q8MY8 PRELIMINARY; PRT; 504 AA.
AC Q8MY8;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE SD12535p.
GN CYP6A9 OR CG10246.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY089375; AAL90113.1; -
DR Flybase; FBgn0038532; CG14322.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR Repeat; WD repeat.
SQ SEQUENCE 1196 AA; 133061 MW; 9ADBA0269C8DEB47 CRC64;

Query Match 15.5%; Score 83.5; DB 5; Length 1196;
Best Local Similarity 30.1%; Pred. No. 0.62;
Matches 28; Conservative 16; Mismatches 32; Indels 17; Gaps 4;

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 2003, 13:55:21; Search time 23 Seconds

(without alignments)  
202.419 Million cell updates/sec

Title: US-09-818-143-22

Perfect score: 539

Sequence: 1 MEALANVPRKSRFPDAG.....RLKVGPRQASQRKLETGIC 99

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	14.7	345	1	CEBB_HUMAN
2	79.5	14.7	998	1	RRPO_BOOLV
3	78.5	14.6	998	1	RRPO_FHV
4	78	14.5	504	1	C6A9_DRONE
5	78	14.5	504	1	C6A1_DRONE
6	78	14.5	825	1	SES_RAT
7	77	14.3	348	1	CEBB_BOVIN
8	75.5	14.0	998	1	RRPO_BBV
9	75	13.9	305	1	GSB2_MOUSE
10	74	13.7	119	1	VTU3_DRONE
11	72	13.4	137	1	VIT_STRPU
12	72	13.4	740	1	POLB_MAIZE
13	70	13.0	2003	1	NTC4_HUMAN
14	69.5	12.9	506	1	C6A2_DRONE
15	69	12.8	323	1	CG96_HUMAN
16	69	12.8	569	1	SILF_MOUSE
17	68	12.6	433	1	FUT4_RAT
18	68	12.6	3726	1	ABF1_MOUSE
19	67.5	12.5	627	1	ACHA_HUMAN
20	67	12.4	597	1	NR41_RAT
21	67	12.4	1235	1	IRS1_RAT
22	66.5	12.3	95	1	Y95_ADE02
23	66.5	12.3	506	1	C6A8_DRONE
24	66.5	12.3	1335	1	XDR_MOUSE
25	66	12.2	507	1	CGA1_DRONE
26	65.5	12.2	1786	1	UVRA_CHLTR
27	65	12.1	172	1	CD42_RAT
28	65	12.1	559	1	DACA_HUMAN
29	65	12.1	606	1	E75_METEN
30	64.5	12.0	457	1	Z185_HUMAN
31	64	11.9	907	1	GP3_HUMAN
32	64	11.9	3703	1	ABF1_HUMAN
33	63.5	11.8	591	1	PAK4_HUMAN

34	63.5	11.8	813	1	YTOJ_CAEBL	Q19673 caenorhabdi
35	63	11.7	222	1	HXC5_HUMAN	Q00444 homo sapien
36	63	11.7	222	1	HXC5_MOUSE	P32043 mus musculus
37	63	11.7	387	1	GT4R_HUMAN	Q9N883 homo sapien
38	63	11.7	390	1	VE2_CRPVK	P03121 cottontail
39	63	11.7	616	1	VGF_HUMAN	O15240 homo sapien
40	63	11.7	670	1	SYN1_MOUSE	O88935 mus musculus
41	63	11.7	704	1	SYN1_RAT	P09951 rattus norv
42	62.5	11.6	406	1	PLG2_HUMAN	Q9BRQ0 homo sapien
43	62.5	11.6	479	1	KLFA_MOUSE	O89091 mus musculus
44	62.5	11.6	598	1	NR41_CANFA	P51666 canis famil
45	62.5	11.6	686	1	DL4_MOUSE	Q9J171 mus musculus

#### ALIGNMENTS

RESULT 1  
CEBB\_HUMAN  
ID CEBB\_HUMAN STANDARD; PRT: 345 AA.  
AC P17676; Q9H425;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE CCAAT/enhancer binding protein beta (C/EBP beta) (Nuclear factor NF-IL6) (Transcription factor 5).  
DE NF-IL6)  
GN CEBPB OR TCF5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCB:TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=90269225; PubMed=2112087;  
RA Akira S., Ishihara H., Sugita T., Tanabe O., Kinoshita S., Nishio Y., Nakajima T., Hirano T., Kishimoto T.  
RT "A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family."  
RL EMBL J. 9:1897-1906(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavridis G., Almeida J.P., Babbage A.K., Baggeley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Lehesvallo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.,  
RT "The DNA sequence and comparative analysis of human chromosome 20."  
RL Nature 414:865-871(2001).  
CC -!- FUNCTION: IMPORTANT TRANSCRIPTIONAL ACTIVATOR IN THE REGULATION OF GENES INVOLVED IN IMMUNE AND INFLAMMATORY RESPONSES. SPECIFICALLY BINDS TO AN IL-1 RESPONSE ELEMENT IN THE IL-6 GENE. NF-IL6 ALSO

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nodaviridae;
OC Alphadnavirus.
OX NCBI_TaxID=12286;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21351039; PubMed=11457991;
RY Johnson K.N., Johnson K.L., Dasgupta R., Ball L.A.;
RT "Comparisons among the larger genome segments of six nodaviruses and
RT their encoded RNA replicases.";
RL J. Gen. Virol. 82:1855-1865(2001).
CC -1- FUNCTION: Replicates the viral genome which is composed of two RNA
CC segments, RNAl and RNAt.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- SIMILARITY: BELONGS TO THE NODAVIRUSES RNA POLYMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF329080; AKK15751.1; -.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
KW Transferase; RNA-directed RNA polymerase.
SQ SEQUENCE 998 AA; 11120 MW; 7FAFFB4FC7329253 CRC64;
Query Match 14.7%; Score 79.5; DB 1; Length 998;
Best Local Similarity 28.2%; Pred. No. 2.1;
Matches 33; Conservative 12; Mismatches 31; Indels 41; Gaps 7;
QY 7 VNPREK----FREDA-GKESG-----QGFCV-----QAAPQPM--VTG 42
DB 628 INCPAKRFRGRTPGMGVSGSPPTTHNTQYNACVEYALKFEYPDANPETFLSLG 687
QY 43 PCSCSPLGNFSPOKENKACACQWAGPKNPCMYRLKVGRPSQRKLKTGLC 99
DB 688 PKCGDDGLARATIKTNRAKCYG :||| :||| :|||
: || : || : ||| :||| :|||
-LELKV-----EKINPEVGLC 727
RESULT 3
RRPO_FHV STANDARD; PRT; 998 AA..
ID RRPO_FHV AC Q6629;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (RDPr) (RNA replicase)
DE (Protein A).
OS Flock house virus (FHV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nodaviridae;
OX Alphadnavirus.
OX NCBI_TaxID=12287;
RN [1]
RP SEQUENCE FROM N.A.
RX Dasgupta R.;
RT "Near identity in the polymerase gene of two serologically distinct
RT nodaviruses.";
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Replicates the viral genome which is composed of two RNA
CC segments, RNAl and RNAt.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- SIMILARITY: BELONGS TO THE NODAVIRUSES RNA POLYMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 29, 2003, 13:55:22 ; Search time 40 Seconds  
(without alignments)  
238.017 Million cell updates/sec

Title: US-09-818-143-22

Perfect score: 539

Sequence: 1 MEALNVNFPKSRPEDAG.....RLKVGQPASQKLTETGLC 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*  
1: piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	539	100.0	99	2 T17268	hypothetical prote
2	86.5	16.0	557	2 T1631	hypothetical prote
3	79.5	14.7	345	2 S12788	transcription fact
4	78.5	14.6	998	2 S41397	protein A - flock
5	78	14.5	825	2 Jc4163	DNA-binding protei
6	75.5	14.0	998	1 QBBB81	B1 protein - black
7	75	13.9	305	2 I57039	genomic screen hom
8	74	13.7	104	2 A22706	vitelline membrane
9	72	13.4	137	2 S00696	vitellogenin precu
10	72	13.4	740	2 T03925	hypothetical prote
11	72	13.4	2333	2 T28675	alpha-51b immobill
12	72	13.4	2333	2 T28674	alpha-51b immobill
13	71.5	13.3	401	2 T02380	hypothetical prote
14	70.5	13.1	3942	2 T42730	Bassoon protein -
15	69.5	12.9	507	1 A47198	cytochrome P450 6A
16	69.5	12.9	511	2 S24345	Balbani ring 1 pr
17	68	12.6	751	2 F87789	protein C3466.2 [1
18	67.5	12.5	98	2 F83404	hypothetical prote
19	67.5	12.5	627	2 Jc4021	nicotinic acetylch
20	67	12.4	563	2 J00623	nerve growth facto
21	67	12.4	1335	1 S16948	insulin receptor s
22	66.5	12.3	75	2 S05589	Balbani ring prot
23	66.5	12.3	95	2 A03858	hypothetical prote
24	66.5	12.3	294	2 S13141	hypothetical prote
25	66.5	12.3	494	2 Jc5320	cytochrome P450 mo
26	66.5	12.3	906	2 T28034	hypothetical prote
27	66.5	12.3	1335	1 X0MSDH	xanthine dehydroge
28	66	12.2	108	2 S37199	protein kinase (EC
29	66	12.2	457	2 E88456	protein W03A5.3 [1

30	66	12.2	670	2 T49510	fibroin-3 related
31	66	12.2	1101	2 T20881	hypothetical prote
32	65.5	12.2	372	2 F86189	hypothetical prote
33	65.5	12.2	1786	1 H1527	probable excluducl
34	65	12.1	845	2 T17291	hypothetical prote
35	64.5	12.0	3938	2 T42761	Bassoon protein -
36	64	11.9	402	2 I46053	connexin44 - bovin
37	64	11.9	2783	1 A41948	alpha-fetoprotein
38	63.5	11.8	360	2 S75453	hypothetical prote
39	63.5	11.8	596	2 F75302	ATP-dependent DNA
40	63.5	11.8	649	2 S67787	hypothetical prote
41	63.5	11.8	736	2 T41259	hypothetical prote
42	63.5	11.8	813	2 T21132	hypothetical prote
43	63	11.7	222	2 S20029	homeotic protein H
44	63	11.7	303	2 T28999	hypothetical prote
45	63	11.7	390	1 WZLRRB	E2 protein - cotto

#### ALIGNMENTS

##### RESULT 1

T17268

hypothetical protein DKFZp434N161.1 - human

C:Species: Homo sapiens (man)

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T17268

R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18723

A:Accession: T17268

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-99 <POU>

A:Cross-references: EMBL:AL117489

A:Experimental source: adult testis; clone DKFZp434N161

C:Genetics:

A:Note: DKFZp434N161.1

Query Match 100.0% Score 539; DB 2; Length 99;

Best Local Similarity 100.0%; Pred. No. 2.6e-47;

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEALNVNFPKSRPEDAGKESGGGFCVPAARQTNMTGPGSCSSPGLQNFSPQRKN' 60

Db 1 MEALNVNFPKSRPEDAGKESGGGFCVPAARQTNMTGPGSCSSPGLQNFSPQRKN 60

Qy 61 RACACWQVAGPAPKPNMPCVRLKVGPRQASQKLTETGLC 99

Db 61 RACACWQVAGPAPKPNMPCVRLKVGPRQASQKLTETGLC 99

##### RESULT 2

T41631

hypothetical protein SPCC830.05c - fission yeast (Schizosaccharomyces pombe).

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T41631

R:McDougal, R.M.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.

submitted to the EMBL Data Library, August 1999

A:Reference number: Z22005

A:Accession: T41631

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-557 <MCD>

A:Cross-references: EMBL:AL109850; PIDN:CA852878.1; GSPDB:GN00068; SPDB:SPCC830.05c

A:Experimental source: strain 972h-; cosmid c830

C:Genetics:

A:Gene: SPDB:SPCC830.05c

A:Map position: 3

A:Introns: 56/3

Query Match 16.0% Score 86.5; DB 2; Length 557;

```

Best Local Similarity 30.4%; Pred. No. 0.42;
Matches 28; Conservative 18; Mismatches 33; Indels 13; Gaps 6;

QY 6 NVNPRKSFRR--PEDAGKES-QSGGFCVPAARQTMVT--GPSGSP--GLQNF-----53
Db 402 HNYPDSPYRIIPIYSGKTKQSHNLSIPSTPLSDNGTPTSTPHSSLSNFTCD 461
QY 54 SPQRKNRACACWQAGPAPKPMCVRLKVR 85
Db 462 SLSPSSNSLYGYSTL-LHPRNPICVRORIGR 492

RESULT 3
S12788
transcription factor NF-IL6 - human
N:Alternate names: nuclear factor NF-IL6
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 29-Sep-1999
C:Accession: S12788
;Akira, S.; Ishihara, H.; Sugita, T.; Tanabe, O.; Kinoshita, S.; Nishio, Y.; Nakajima, T.
A:Title: A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family.
A:Reference number: S12788; MUID:90269225; PMID:2112087
A:Accession: S12788
A:Molecule type: DNA
A:Residues: 1-345 <AKI>
A:Cross-references: EMBL:X52560; NID:g35035; PIDN:CAA36794.1; PID:g35036
C:Genetics:
A:Gene: GDB:CEBPB; LAP; TCF5; CRP2; NFIL6; IL6BP
A:Cross-references: GDB:126373; OMIM:189965
A:Map position: 20q12-20q13.1
C:Superfamily: CCAAT/enhancer-binding protein alpha
C:Keywords: DNA binding; transcription factor

Query Match 14.7%; Score 79.5; DB 2; Length 345;
Best Local Similarity 26.2%; Pred. No. 1.3;
Matches 32; Conservative 10; Mismatches 39; Indels 41; Gaps 5;

QY 10 PRKSFPEPDAGKESGSG--GF-----CYPAAQPOTMTGTPSCSPGLQNFSP 55
Db 181 PADCKRKEAGAPGGAGAGFPVALRYLGYQVPSGSSLSSTSSSPFGTP--SP 238
QY 56 QRKNRACACWQAGPAP-----KNPMCVRLKVRPQASQK 92
Db 239 ADAKAPTACIAGAGAPSPQVSKAKTKVDKSHDEYKIRRNRIAVR--KSRQAKMRN 296
QY 93 LK 94
b 297 LE 298

RESULT 4
S41397
Protein A - flock house virus
C:Species: flock house virus
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
C:Accession: S41397
R:Dasgupta, R.
submitted to the EMBL Data Library, January 1994
A:Description: Near identity in the polymerase gene of two serologically distinct nodaviruses.
A:Reference number: S41397
A:Accession: S41397
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-998 <DAS>
A:Cross-references: EMBL:X77156; NID:g450500; PIDN:CAA54399.1; PID:g450501
C:Superfamily: black beetle virus B1 protein

Query Match 14.6%; Score 78.5; DB 2; Length 998;
Best Local Similarity 29.1%; Pred. No. 4.6;
Matches 34; Conservative 10; Mismatches 32; Indels 41; Gaps 7;

QY 7 VNFPRKS-----FRPE-DAGKESGS-----QGQFCV-----PAARPQTM--VTG 42

```

```

Db 628 INCPAKRFGFFRYEFGVGKSGSTTPHTQTCGVEFTALTFFHPDAEPEDLRLIG 687
QY 43 PSCSSPGLQNFSPQRKNRACACWQAGPAPKPMCVRLKVRPQASQKRLKETGLC 99
Db 688 PKCGDDGLSRALIOKSIINRAKCFG-----LELV-----ERYNFEIGLC 727

```

```

RESULT 5
JC4163
DNA-binding protein 5E5 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 20-Jun-2000
C:Accession: JC4163; PC4040
R:Suzuki, E.; Kojima, N.; Yoshimura, K.; Uyenura, K.; Obata, K.; Akagawa, K.
J. Biochem. 118, 122-128, 1995
A:Title: Cloning and sequence analysis of cDNA for a possible DNA-binding protein 5E5
A:Reference number: JC4163; MUID:96015159; PMID:8537300
A:Accession: JC4163
A:Molecule type: mRNA
A:Residues: 1-825 <SUZ>
A:Cross-references: DDBJ:D37934; NID:g531260; PIDN:BAA07153.1; PID:g531261
A:Experimental source: brain
A:Accession: PC4040
A:Molecule type: protein
A:Residues: 230-455 <SUZ>
A:Comment: This protein has an abundance of arginine, a glycine-rich region and a pro
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C:Keywords: nerve; phosphoprotein
F:436-443/Region: nuclear location signal
F:722-731/Region: proline cluster
F:62,258,345,360,404,413,570,613,752,820/Binding site: phosphate (Ser) (covalent)

```

```

Query Match 14.5%; Score 78; DB 2; Length 825;
Best Local Similarity 30.3%; Pred. No. 4.3;
Matches 27; Conservative 13; Mismatches 25; Indels 24; Gaps 4;

QY 19 AGKESGSGGFCVPAARQTMVTGTPSCSPGLQNFSPQRKNRACACWQ-----68
Db 742 AGDPGASRSGRRPARRRGELT-PQRPSP-----FAPQEGRLAESCVDDGAIAPDTDTA 796
QY 69 -----AGPAPKPMCVRLKVRPQASQK 91
Db 797 SGVEPAGPSLSSTMC---QNGRPRSPK 822

```

```

RESULT 6
QB8BI
B1 protein - black beetle virus
C:Species: black beetle virus
C:Date: 30-Sep-1992 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S78471; S28728; S28729; A23243
R:Dasgupta, R.
submitted to the EMBL Data Library, July 1986
A:Reference number: S78471
A:Accession: S78471
A:Molecule type: genomic RNA
A:Residues: 1-998 <DAS>
A:Cross-references: EMBL:X02396; NID:g60679; PIDN:CAA26238.1; PID:g60680
R:Dasgupta, R.; Ghosh, A.; Kaesberg, P.
J. Mol. Biol. 182, 183-189, 1985
A:Title: Structure of the black beetle virus genome and its functional implications.
A:Reference number: S28728; MUID:85210903; PMID:3839022
A:Accession: S28728
A:Molecule type: genomic RNA
A:Residues: 1-883; AALRRPPTNRYQC' <DAS>
A:Cross-references: EMBL:X02396
A:Accession: S28729
A:Molecule type: genomic RNA
A:Residues: 897-998 <DA2>
A:Cross-references: EMBL:X02396
R:Guarino, L.A.; Ghosh, A.; Dasgupta, R.; Kaesberg, P.
Virology 139, 199-203, 1984

```